CardioCNN: Gamification of counterfactual image transplants

Hakan Lane[1]
hlane@uni-mainz.de
Michal Valko[2]
misko.valko007@gmail.com.
Stefan Kramer[1]
www.datamining.informatik.uni-mainz.de

[1] Data Mining
Johannes Gutenberg University
Mainz, Germany

[2] Theological Institute
Catholic University in Ruzomberok
Ruzomberok, Slovakia

Abstract

Supervised image classification with deep learning holds significant potential of automated early detection of numerous medical conditions. However, the deployment in clinics lags behind the technical potential. There is a widespread need to create services that can interact better with patients and doctors to explain the medical predictions. One, crucial, feature enhancing this interaction is the ability to make incremental changes to the input and observe the results. This paper presents an approach whereby users can translate patches from a source image to a target through a web user interface. While the prediction is updated through applying the trained network on the merged semifactual scan. The procedure is framed as a game with the stated goal of an upward transition with minimal effort, i.e. converting the target picture with an unhealthy label into a counterfactual healthy scan by translating patches from a source image of a healthy individual. Success is measured by minimising the translated fraction, i.e. making the intervention as non-invasive as possible. Our method relies on superpixel segments, an interpolating morphing flood fill method to smoothly translate patches between images, and a convolutional neural network (CNN). The first case study was performed on chest X-rays for identification of cardiomegaly (CM) showing that the prediction can be reversed by copying only a few % of the image when the source and target areas are selected carefully. This enables a better understanding of the clinical domain and the deep learning methods as well as a platform for raising health awareness.

1 Introduction

Supervised learning for image classification has reached a mature stage in heart disease prediction, being applied to cardiovascular scanning outputs to detect a variety of conditions [6]. Common modalities include magnetic resonance imaging (MRI), echocardiograms and chest X-rays, and methods include convolutional neural networks (CNN), Transfer learning, Long short term memory, Generative adversarial networks (GAN), Autoencoders and hybrids [18]. With accuracies surpassing 90 %, these systems arguably outperform human capacity in both performance and efficiency [7]. The use of Explainable Artificial Intelligence (XAI) in medical imaging aims to provide more transparent and interpretable AI solutions to assist healthcare professionals [12]. Despite this progress, the deployment of deep learning in clinical imaging settings lags the technical potential, inhibited due to lack of understanding and trust between the domain experts and developers. To address the mentioned gap, the research has increasingly focused on XAI techniques, such as Grad-CAM, heatmaps and counterfactual visualisations. These techniques have been applied not only for prediction but also to simulation of treatment effects and even surgical outcomes [36].

We propose a gamified counterfactual technique that allows the user to test the effect of a virtual transplant of a selected area from a healthy person's scan onto a similarly arbitrarily chosen segment of a diagnosed image. The method combines segmentation, morphing image mergers and a CNN classifier to create a competitive framework aimed at reversing the unhealthy state by translating patches between the images. Our goal is to enhance understanding, raise awareness and motivate end users.

2 Previous Work

2.1 Cardiovascular Image Classification

Cardiovascular image classification is a pivotal area in medical imaging, significantly contributing to the diagnosis and treatment of cardiovascular diseases (CVDs). Recent advancements in machine learning and deep learning have propelled significant progress in the field, introducing novel methodologies and tools for more accurate and efficient analysis. Deep learning, particularly CNNs, have emerged as the dominant technique for cardiac image segmentation. This process is essential for identifying anatomical structures such as ventricles, atria, and coronary arteries. The capability of CNNs to learn hierarchical features from imaging data enables highly precise segmentation and analysis [10]. Another key component of cardiovascular image classification is the analysis of electrocardiogram (ECG) images. Lightweight CNNs augmented with attention modules have demonstrated high accuracy in detecting various abnormalities. These models can effectively categorise ECG images into classes such as abnormal heartbeat, myocardial infarction, and normal ECG, thus facilitating timely and accurate diagnosis [29].

The advancement of software tools and computational frameworks has further supported cardiovascular image analysis. These innovations integrate sophisticated image processing techniques and machine learning algorithms, enhancing both the accuracy and efficiency of image classification. Such innovations are essential for translating research advancements into clinical practice [34]. Despite these advancements, numerous challenges persist in the field of cardiovascular image classification. The scarcity of labelled data, the need for model generalisability across different imaging modalities, and the interpretability of deep learning models are key obstacles. Addressing these challenges is crucial for improving the robustness and applicability of cardiovascular image classification systems. Future research should focus on developing methods to overcome these obstacles, ensuring that the full potential of these technologies can be realised in clinical settings [10].

2.2 Counterfactual visualisation

Counterfactual refers to hypothetical scenario (what the image would look like with the attribute increased or decreased) often described as "What-if" scenarios. Counterfactual techniques have emerged as a powerful approach in deep learning, enabling the extraction of causal insights from complex models [22]. In the domain of cardiovascular imaging, these methods provide a framework to analyse "what-if" scenarios, assisting clinicians in better understanding disease progression, treatment impacts, and patient outcomes. In medical imaging, counterfactual analysis can help predict outcomes under alternative treatment regimes or understand how changes in patient characteristics could affect disease progression. Traditional models are often limited to correlational analysis, whereas counterfactual approaches incorporate causal reasoning, thus offering a more nuanced view of patient data [33].

Counterfactual techniques can enhance the predictive performance of deep learning models by assessing how modifications in imaging features influence the prognosis of cardiovascular diseases. For instance, by the integration of counterfactual reasoning with convolutional neural networks (CNNs), have demonstrated improved risk stratification in patients with coronary artery disease [2]. By employing counterfactual analysis, clinicians can simulate alternative therapeutic interventions, assisting in clinical decision-making. For example, Nguyen *et al*. (2020) [23] utilised counterfactual models to evaluate treatment effects for patients with various medical diseases, revealing valuable insights into optimal patient management strategies based on patient-specific imaging data. The authors also uncover significant potential in advancing medical decision-making and diagnosing processes [23].

Counterfactual techniques enable for the modelling of disease trajectories, helping to elucidate how specific factors contribute to the progression of cardiovascular conditions. Empirical studies have employed these methods such as risk prediction model offering probabilistic model for diagnosing cardiovascular disease [8]. Although counterfactual techniques offer substantial benefits, they also present several challenges. The development of accurate counterfactual models requires careful consideration of confounding variables and the selection of appropriate reference scenarios. Additionally, robustness to model misspecifications and data sparsity are essential issues that need to be addressed for reliable outcomes [8] [25].

Furthermore, the interpretability of models remains a crucial aspect in the medical domain. Researchers must ensure that counterfactual analyses are transparently communicated to clinicians, fostering trust and facilitating clinical adoption [13].

2.3 Gamification for increased medical awareness

Gamification refers to the application of elements such as leaderboards. The primary goal of gamification is to stimulate of structural and trait-based competitiveness (engagement, performance growth or informal learning environments) [4]. For educational purposes, gamification have been found to increase student engagement and improve learning outcomes compared to traditional learning methods [20]. In terms of fitness apps, incommensurate elements (such as likes) have been found to increase intrinsic motivation more effectively compared to commensurate elements (points, score) [11].

Gamification has evolved beyond entertainment and is now widely utilised for education purposes as well. The educational processes can be enhanced by integration of virtual environment. The theoretical framework was also empirically investigated in various contexts such as radiology [3]. The robust potential of gamification for medical students as end users offers prospects not only for enhancing the learning process itself, but for improving training in decision-making processes and decision-making awareness as part of diagnose processes. [16].

Empirical studies further explore novel VR-based gaming methods. The findings show promising results not only with diagnosing from medical images VR simulator as effective learning tool that facilitate the developments of technical skills among future physicians [14].

Medical imaging games can be broadly classified a) Games utilised for educational purposes [3];[16], and b) Games utilised for better diagnosing processes [3]; [16];[14].

These innovations enhance the learning experience for medical professionals and increase the trust on AI-assisted diagnoses [26]. In the case of radiographs, Winkel *et al.* [38] have presented their RapRad game, where the main objective of the game was an engaging learning environment combining elements of gaming and medical training. A total of 195 game levels covered various aspects of the presence or absence of pneumothorax, with a point based scoring model to ensure intrinsic motivation from the users [11]. Another radiology Pocket Game App for training was developed by Prasath who uses a simple interface where user taps on the correct answer (green correct answer, red incorrect) [27].

3 Method

The translation technique works in three steps, as illustrated in Figure 1.

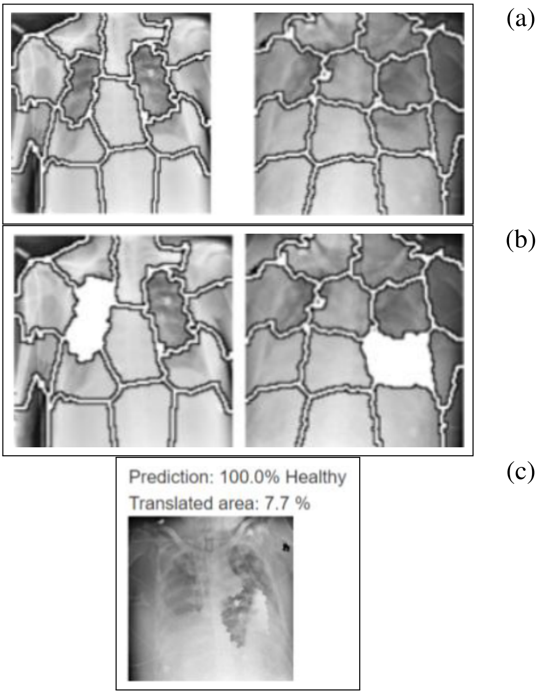


Figure 1: Description of game steps a) original images: right healthy, left cardiomegaly diagnosis b) area selection: right healthy (participants choose area which will be transplanted), left cardiomegaly (participants choose where to insert translated area from right picture) to make patient completely healthy c) result: after participants click on transplant, the result will appear (percentage of healthy or cardiomegaly prediction and percentage of translated area statement from healthy inserted to cardiomegaly X-ray).

3.1 Segmentation

In the initial step, the image is segmented into a user-assigned number of areas via the superpixels method [28]. Superpixels are groups of pixels that share similar characteristics, such as colour or texture, and are used to simplify the segmentation into patches [32]. They aggregate perceptually similar pixels into meaningful clusters, reducing the complexity of image data and serving as a preprocessing step for many image segmentation tasks [37]. The segmentation in the game algorithm uses Simple Linear Iterative Clustering (SLIC) based on k-means clustering to group pixels based on colour similarity and spatial proximity [1]. Figure 2 illustrates the effect of SLIC with varying parameters governing the size of the segments.

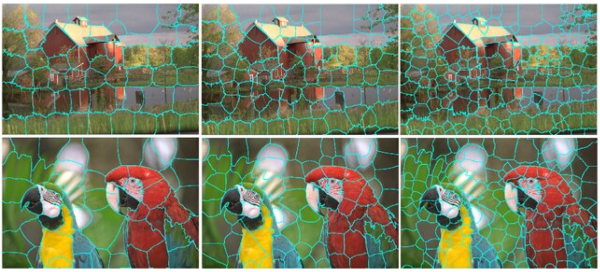


Figure 2: SLIC Superpixel segmentation with different number of superpixels

The implementation uses the OpenImageR package [21] with a compactness of 20 (fixed), while the number of superpixels can be selected by the user in the range from 6 to 30.

3.2 Morph insertion

The MIMICRI method, introduced by Guo *et al.* in 2024 [15], is a Python library [1] designed to provide domain-centered counterfactual explanations for cardiovascular image classification models. It uses a recombination method called Morphmix to fill selected segments in the target with pixels coming from the chosen segment in the target. A heuristic technique to align the centroids of the segments designated for replacement was employed. From these centroids, we apply a flood-fill algorithm to copy pixels from the original image into the target location. The result is a recombined image with the selected segments replaced appropriately, as displayed on an MRI image in Figure 3.

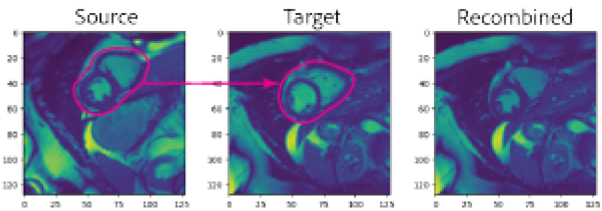


Figure 3: MIMICRI insertion example. Copied with permission from [15]

[1] https://github.com/IBM/mimicri

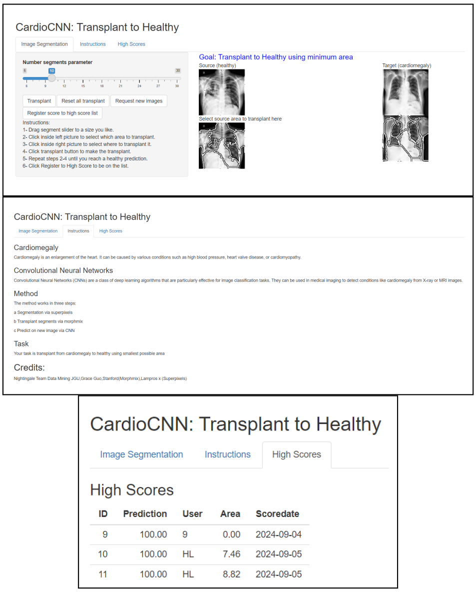
3.3 Prediction

The method is agnostic in regard to the underlying image classification model. In the presented trial, a convolutional neural network was employed, with an architecture of Convolution - Activation - Convolution - Leaky Rectified Linear Unit - Batch Normalization - Max Pooling - Dropout - Flatten - Dense - Activation - Dropout - Dense - Activation.

4 Game presentation

4.1 User Interface

The user interface runs as an R shiny web based app connected to a MySQL database with three tabs as presented in Figure 4



(a)
(b)
(c)

Figure 4: User Interface: a) Game page b) Instructions c) High score list

4.2 Game features

The goal of the game is to transplant healthy tissue into an image with cardiomegaly (CM). The main objective lies in achieving an upward transition to at least 50% prediction of healthy with the smallest possible minimal intervention (minimal transplantation area). Successful users can store their results on a high scoreboard to improve engagement [19]. The progress scoring is provided immediately showing the percentage of the image area, a score that can be registered to empower intrinsic motivation [4].
Figure 5 illustrates a successful process.

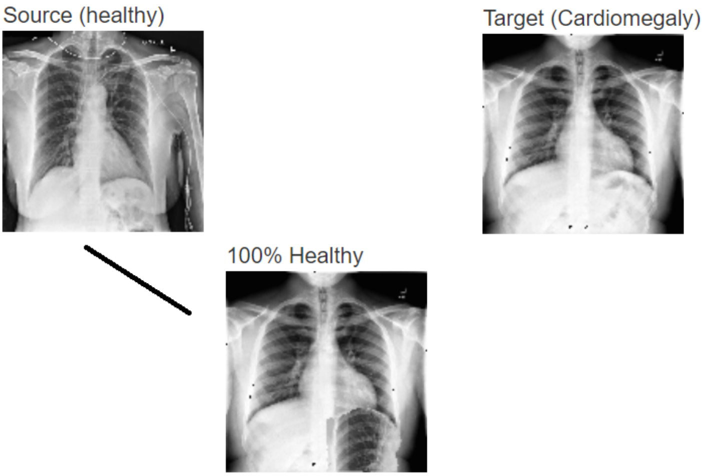


Figure 5: Example successful translation

4.3 Software

R Shiny [9] was used to create the interface, together with the packages OpenImageR [21] (superpixels) and shinyjs [6] for the graphical interaction. Deep Learning models are created using the keras package [17] in a Python virtual environment. File and Database I/O was handled by the png [35] and RMySQL [24] packages.

4.4 User Explainability Tests

The service was tested by a convenience sample recruited via various online survey platforms and measured whether testing the game could help users understand cardiomegaly better and where in a scan they should look for traces of the condition. They were divided into three groups as specified by their different support visuals of a) seeing a Grad-CAM heatmap of a CNN cardiomegaly prediction, b) Control (no help), and c) being asked to play the game for some time. They were ask to select a region in Figure 6 to heal. Psytoolkit [30, 31] was used.

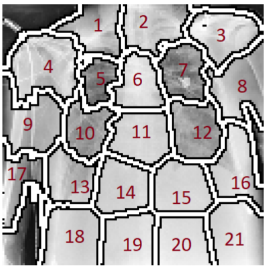
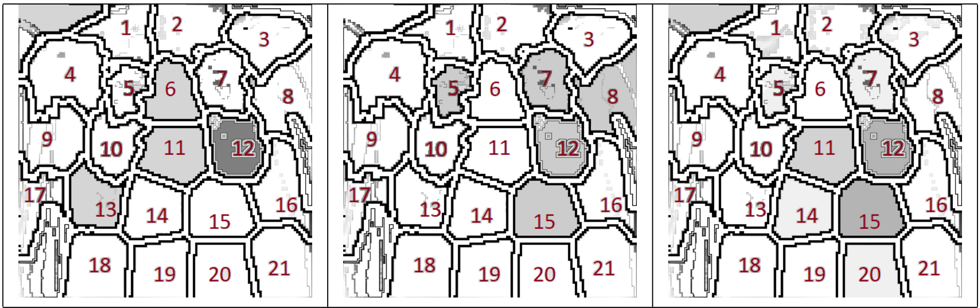


Figure 6: Segmented image for user survey

5 Results

Figure 7 illustrates the game-user choice heatmaps across three experimental conditions. The heatmaps reflect the areas selected by participants in each group when they were prompted to identify key regions.



a) Control group	b) Game play	c) Grad-CAM Heatmap

Figure 7: User choice heatmaps based on cue

The intervened groups demonstrate a slight shift to right and down.

6 Discussion

The preliminary test findings indicate that the interventions caused a shift in user selection towards critical areas beneath the ribs where the CM heart expands. This observation aligns with predicted anatomical changes during CM diagnosing, arguing that the game successfully altered respondent grasp of CM structure, much like previous medical student scorecards [16]. GRAD-CAM was even more efficient.

7 Conclusions

The presented method highlights a potential pathway for integrating gamification into medical AI for better understanding of the pathological domain as well as the deep learning mechanisms. User tests indicate an early tendency to select key regions when asked about where to act against cardiomegaly, indicating that the gamified service can act to train people in domain knowledge. Considerable future work is needed, particularly in user testing, to determine the most effective ways to present the service, the optimal scanning systems for its use, and its integration with other clinical systems.

8 Acknowledgments

The project was funded by the Federal Ministry for Education and Research (BMBF) with the grant number 03ZU120212. The results images were created via the online tool Lunapic. [2] Grace Guo from Harvard University provided code for the Morph insertion function based on the Mimicri code [15].

[2] www.lunapic.com

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012. doi: 10.1109/TPAMI.2012.120.

[2] Rafael Adolf, Nejva Nano, Alessa Chami, Claudio E von Schacky, Albrecht Will, Eva Hendrich, Stefan A Martinoff, and Martin Hadamitzky. Convolutional neural networks on risk stratification of patients with suspected coronary artery disease undergoing coronary computed tomography angiography. *The International Journal of Cardiovascular Imaging*, 39(6):1209–1216, 2023. doi: 10.1007/s10554-023-02824-y.

[3] Pedro Aguado-Linares and Francisco Sendra-Portero. Gamification: Basic concepts and applications in radiology. *Radiología*, 65(2):122–132, 2023. doi: 10.1016/j.rxeng.2022.10.014.

[4] Laura Amo, Ruochen Liao, Rajiv Kishore, and Hejamadi R Rao. Effects of structural and trait competitiveness stimulated by points and leaderboards on user engagement and performance growth: A natural experiment with gamification in an informal learning environment. *European Journal of Information Systems*, 29(6):704–730, 2020. doi: 10.1080/0960085X.2020.1808540.

[5] Dean Attali. shinyjs: Easily improve the user experience of your shiny apps in seconds, 2015. URL http://dx.doi.org/10.32614/CRAN.package.shinyjs.

[6] Rubén G Barriada and David Masip. An overview of deep-learning-based methods for cardiovascular risk assessment with retinal images. *Diagnostics*, 13(1), 2023. doi: 10.3390/diagnostics13010068.

[7] Sanghita Barui, Parikshit Sanyal, K S Rajmohan, Ajay Malik, and Sharmila Dudani. Perception without preconception: comparison between the human and machine learner in recognition of tissues from histological sections. *Scientific Reports*, 12(1), 2022. doi: 10.1038/s41598-022-20012-1.

[8] Christopher B Boyer, Issa J Dahabreh, and Jon A Steingrimsson. Assessing model performance for counterfactual predictions. *arXiv preprint arXiv:2308.13026*, 2023.

[9] Winston Chang, Joe Cheng, J J Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. shiny: Web application framework for r, 2012.

[10] Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine*, 7:25, 2020. doi: 10.3389/fcvm.2020.00025.

[11] Wenting Feng, Rungting Tu, and Peishan Hsieh. Can gamification increases consumers' engagement in fitness apps? the moderating role of commensurability of the game elements. *Journal of Retailing and Consumer Services*, 57:102229, 2020. doi: 10.1016/j.jretconser.2020.102229.

[12] Miguel Fontes, João Dallyson Sousa De Almeida, and António Cunha. Application of example-based explainable artificial intelligence (xai) for analysis and interpretation of medical imaging: a systematic review. *IEEE Access*, 12:26419–26427, 2024. doi: 10.1109/ACCESS.2024.3367606.

[13] Maria Frasca, Davide La Torre, Gabriella Pravettoni, and Ilaria Cutica. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discover Artificial Intelligence*, 4(1):15, 2024. doi: 10.1007/s44163-024-00114-7.

[14] Therese Gunn, Lee Jones, Pete Bridge, Pam Rowntree, and Lisa Nissen. The use of virtual reality simulation to improve technical skill in the undergraduate medical imaging student. *Interactive Learning Environments*, 26(5):613–620, 2018. doi: 10.1080/10494820.2017.1374981.

[15] Grace Guo, Lifu Deng, Animesh Tandon, Alex Endert, and Bum Chul Kwon. Mimicri: Towards domain-centered counterfactual explanations of cardiovascular image classification models. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1861–1874, 2024. doi: 10.1145/3630106.3659011.

[16] Kosuke Ishizuka, Kiyoshi Shikino, Hajme Kasai, Yoji Hoshina, Saito Miura, Tomoko Tsukamoto, Kazuyo Yamauchi, Shoichi Ito, and Masatomi Ikusaka. The influence of gamification on medical students' diagnostic decision making and awareness of medical cost: a mixed-method study. *BMC Medical Education*, 23(1):813, 2023. doi: 10.1186/s12909-023-04808-x.

[17] Tomasz Kalinowski, J J Allaire, and François Chollet. keras3: R interface to "keras", 2024.

[18] Rakesh Kumar, Pooja Kumbharkar, Sandeep Vanam, and Sanjeev Sharma. Medical images classification using deep learning: a survey. *Multimedia Tools and Applications*, 83(7):19683–19728, 2023. doi: 10.1007/s11042-023-15576-7.

[19] Rocio Lorenzo-Alvarez, Teodoro Rudolphi-Solero, Miguel J Ruiz-Gomez, and Francisco Sendra-Portero. Game-based learning in virtual worlds: a multiuser online game for medical undergraduate radiology education within second life. *Anatomical sciences education*, 13(5):602–617, 2020. doi: 10.1002/ase.1927.

[20] John Aries I Malahito and Maria Ana T Quimbo. Creating g-class: A gamified learning environment for freshman students. *E-Learning and Digital Media*, 17(2):94–110, 2020. doi: 10.1177/2042753019899805.

[21] Lampros Mouselimis. Openimager: An image processing toolkit, July 2016. URL http://dx.doi.org/10.32614/CRAN.package.OpenImageR.

[22] Supriya Nagesh, Nina Mishra, Yonatan Naamad, James M Rehg, Mehul A Shah, and Alexei Wagner. Explaining a machine learning decision to physicians via counterfactuals. In *Conference on Health, Inference, and Learning*, pages 556–577, 2023. URL https://proceedings.mlr.press/v209/nagesh23a.html.

[23] Tri-Long Nguyen, Gary S Collins, Paul Landais, and Yannick Le Manach. Counterfactual clinical prediction models could help to infer individualized treatment effects in randomized controlled trials—an illustration with the international stroke trial. *Journal of clinical epidemiology*, 125:47–56, 2020. doi: 10.1016/j.jclinepi.2020.05.022.

[24] Jeroen Ooms, David James, Saikat DebRoy, Hadley Wickham, and Jeffrey Horner. Rmysql: Database interface and "mysql" driver for r, 2000.

[25] Martin Pawelczyk, Klaus Broelemann, and Gjergji Kasneci. Learning model-agnostic counterfactual explanations for tabular data. In *Proceedings of the web conference 2020*, pages 3126–3132, 2020. doi: 10.1145/3366423.3380087.

[26] Enrica Pesare, Teresa Roselli, Nicola Corriero, and Veronica Rossano. Game-based learning and gamification to promote engagement and motivation in medical learning contexts. *Smart Learning Environments*, 3:1–21, 2016. doi: 10.1186/s40561-016-0028-0.

[27] V B Surya Prasath. App review series: Radiology pocket game. *Journal of Digital Imaging*, 30:127–129, 2017. doi: 10.1007/s10278-016-9924-7.

[28] Ren and Malik. Learning a classification model for segmentation. In *Proceedings Ninth IEEE International Conference on Computer Vision*, 2003. doi: 10.1109/iccv.2003.1238308.

[29] Tariq Sadad, Mejdl Safran, Inayat Khan, Sultan Alfarhood, Razaullah Khan, and Imran Ashraf. Efficient classification of ecg images using a lightweight cnn with attention module and iot. *Sensors*, 23(18):7697, 2023. doi: 10.3390/s23187697.

[30] Gijsbert Stoet. Psytoolkit: A software package for programming psychological experiments using linux. *Behavior Research Methods*, 42(4):1096–1104, 2010. doi: 10.3758/brm.42.4.1096.

[31] Gijsbert Stoet. Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1):24–31, 2016. doi: 10.1177/0098628316677643.

[32] Subhashree Subudhi, Ram Narayan Patro, Pradyut Kumar Biswal, and Fabio Dell'Acqua. A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5015–5035, 2021. doi: 10.1109/JSTARS.2021.3076005.

[33] Toygar Tanyel, Serkan Ayvaz, and Bilgin Keserci. Beyond known reality: Exploiting counterfactual explanations for medical research. *arXiv preprint arXiv:2307.02131*, 2023.

[34] Adam Updegrove, Nathan M Wilson, Jameson Merkow, Hongzhi Lan, Alison L Marsden, and Shawn C Shadden. Simvascular: an open source pipeline for cardiovascular simulation. *Annals of biomedical engineering*, 45:525–541, 2017. doi: 10.1007/s10439-016-1762-8.

[35] Simon Urbanek. png: Read and write png images, 2010.

[36] Bas H M van der Velden, Hugo J Kuijf, Kenneth G A Gilhuijs, and Max A Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, 79:102470, 2022. doi: https://doi.org/10.1016/j.media.2022.102470.

[37] Xueqian Wang, Gang Li, Antonio Plaza, and You He. Revisiting slic: Fast superpixel segmentation of marine sar images using density features. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022. doi: 10.1109/TGRS.2022.3142068.

[38] David J Winkel, Philipp Brantner, Jonas Lutz, Safak Korkut, Sebastian Linxen, and Tobias J Heye. Gamification of electronic learning in radiology education to improve diagnostic confidence and reduce error rates. *American Journal of Roentgenology*, 214 (3):618–623, 2020. doi: 10.2214/AJR.19.22087.